How to calculate p-values

Statistics and Big Data

Niccolò Salvini, PhD

UCSC

Academic Year 2025-2026

Course: Statistics and Big Data

Overview

1. What Are P-Values?
2. The Practical Steps of Hypothesis Testing
3. Significance vs. Importance
4. Summary and Exercises

Defining P-Values

We've discussed rejecting H_0 if our data is "unlikely". The p-value gives us a precise measure of that "unlikeliness".

Definition

The **p-value** is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from our sample data, *assuming the null hypothesis (H_0) is true*.

Defining P-Values

We've discussed rejecting H_0 if our data is "unlikely". The p-value gives us a precise measure of that "unlikeliness".

Definition

The **p-value** is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from our sample data, *assuming the null hypothesis* (H_0) *is true*.

$$p\text{-value} = P(\text{observed data or more extreme} | H_0 \text{ is true})$$

Defining P-Values

We've discussed rejecting H_0 if our data is "unlikely". The p-value gives us a precise measure of that "unlikeliness".

Definition

The **p-value** is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from our sample data, *assuming the null hypothesis (H_0) is true*.

$$p\text{-value} = P(\text{observed data or more extreme} | H_0 \text{ is true})$$

Interpretation

A small p-value ($\leq \alpha$) means our data is surprising if H_0 is true, so we reject H_0.
A large p-value ($> \alpha$) means our data is not surprising if H_0 is true, so we fail to reject H_0.

The 5-Step Procedure for Hypothesis Testing

Here is a systematic approach to conduct any hypothesis test using the p-value method.

1. **Develop the null (H_0) and alternative (H_1) hypotheses.**

The 5-Step Procedure for Hypothesis Testing

Here is a systematic approach to conduct any hypothesis test using the p-value method.

1. **Develop the null (H_0) and alternative (H_1) hypotheses.**
2. **Specify the level of significance (α).** (e.g., $\alpha = 0.05$)

The 5-Step Procedure for Hypothesis Testing

Here is a systematic approach to conduct any hypothesis test using the p-value method.

1. **Develop the null (H_0) and alternative (H_1) hypotheses.**
2. **Specify the level of significance (α).** (e.g., $\alpha = 0.05$)
3. **Collect sample data and compute the value of the test statistic.** (e.g., z-score or t-score)

The 5-Step Procedure for Hypothesis Testing

Here is a systematic approach to conduct any hypothesis test using the p-value method.

1. **Develop the null (H_0) and alternative (H_1) hypotheses.**
2. **Specify the level of significance (α).** (e.g., $\alpha = 0.05$)
3. **Collect sample data and compute the value of the test statistic.** (e.g., z-score or t-score)
4. **Use the test statistic to compute the p-value.**

The 5-Step Procedure for Hypothesis Testing

Here is a systematic approach to conduct any hypothesis test using the p-value method.

1. **Develop the null (H_0) and alternative (H_1) hypotheses.**
2. **Specify the level of significance (α).** (e.g., $\alpha = 0.05$)
3. **Collect sample data and compute the value of the test statistic.** (e.g., z-score or t-score)
4. **Use the test statistic to compute the p-value.**
5. **Make the decision: Reject H_0 if p-value $\leq \alpha$.**

Choosing the Right Test Statistic

The key question: is the population standard deviation σ known?

Case 1: σ Known

By the Central Limit Theorem, $\bar{m}$ is normally distributed. Use the Z distribution:

$$z = \frac{m - \mu_0}{\sigma/\sqrt{n}}$$

Choosing the Right Test Statistic

The key question: is the population standard deviation σ known?

Case 1: σ Known

By the Central Limit Theorem, $\bar{m}$ is normally distributed. Use the Z distribution:

$$z = \frac{m - \mu_0}{\sigma/\sqrt{n}}$$

Case 2: σ Unknown

Replace σ with the sample standard deviation s. Extra variability $\Rightarrow$ use Student's t with $n - 1$ d.f.:

$$t = \frac{m - \mu_0}{s/\sqrt{n}}$$

Step 0: The Scenario

Problem

An Emergency Medical Service (EMS) has a service goal of a mean response time of **12 minutes or less**.

- A random sample of **40 emergencies** is taken.
- The sample mean response time is $\bar{x} = 13.25$ minutes.
- The population standard deviation is known to be $\sigma = 3.2$ minutes.

The director wants to test if the service goal is being achieved, using a **0.05 level of significance**.

The sample mean (13.25) is higher than the goal (12). Is this difference statistically significant, or could it be due to random sampling variation?

Hypotheses and Significance Level

Step 1. Develop the hypotheses

The director is concerned if the time is **greater** than 12 minutes. This suggests a one-tailed test.

- $H_0 : \mu \leq 12$ (The response goal is being met)
- $H_1 : \mu > 12$ (The response goal is NOT being met)

Hypotheses and Significance Level

Step 1. Develop the hypotheses

The director is concerned if the time is **greater** than 12 minutes. This suggests a one-tailed test.

- $H_0 : \mu \leq 12$ (The response goal is being met)
- $H_1 : \mu > 12$ (The response goal is NOT being met)

Step 2. Specify the level of significance

The problem states to use $\alpha = 0.05$.

Compute the Test Statistic

Step 3. Compute the value of the test statistic

Since the population standard deviation σ is known, we use the z-statistic. The formula measures how many standard errors the sample mean $\bar{x}$ is from the hypothesized population mean μ_0.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Compute the Test Statistic

Step 3. Compute the value of the test statistic

Since the population standard deviation σ is known, we use the z-statistic. The formula measures how many standard errors the sample mean $\bar{x}$ is from the hypothesized population mean μ_0.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Plugging in our values:

$$z = \frac{13.25 - 12}{3.2/\sqrt{40}} = \frac{1.25}{0.506} = 2.47$$

Our sample mean is 2.47 standard errors above the target of 12 minutes. Is that a lot?

Compute p-value and Make Decision

Step 4. Use the test statistic to compute the p-value

The p-value is the probability of getting a z-score of 2.47 or higher, if H_0 were true. We are looking for $P(Z \geq 2.47)$.
We can find this from a standard normal table or software.

$$p\text{-value} = P(Z \geq 2.47) = 1 - P(Z < 2.47) = 1 - 0.9932 = 0.0068$$

Compute p-value and Make Decision

Step 4. Use the test statistic to compute the p-value

The p-value is the probability of getting a z-score of 2.47 or higher, if H_0 were true. We are looking for $P(Z \geq 2.47)$.
We can find this from a standard normal table or software.

$$p\text{-value} = P(Z \geq 2.47) = 1 - P(Z < 2.47) = 1 - 0.9932 = 0.0068$$

Step 5. Make the decision

The rule is: Reject H_0 if p-value $\leq \alpha$.

- Our p-value is 0.0068.
- Our α is 0.05.
- Since $0.0068 \leq 0.05$, we **reject the null hypothesis**.

Conclusion: There is sufficient statistical evidence to infer that the EMS is not

The Importance of Effect Size

While a small p-value indicates a statistically significant difference, it does not convey the *magnitude* or *practical importance* of that difference.

The Importance of Effect Size

While a small p-value indicates a statistically significant difference, it does not convey the *magnitude* or *practical importance* of that difference.

- With a very large sample size (n), even a tiny, meaningless difference can become statistically significant (i.e., have a very small p-value).
- In our EMS example, we rejected H_0. The difference was 1.25 minutes. Is this practically important? That's a question for the EMS director, not for the p-value to answer.

The Importance of Effect Size

While a small p-value indicates a statistically significant difference, it does not convey the *magnitude* or *practical importance* of that difference.

- With a very large sample size (n), even a tiny, meaningless difference can become statistically significant (i.e., have a very small p-value).
- In our EMS example, we rejected H_0. The difference was 1.25 minutes. Is this practically important? That's a question for the EMS director, not for the p-value to answer.

Key Insight

Statistical significance $\neq$ Practical significance. Always report and consider the effect size (e.g., the difference in means) alongside the p-value.

Summary of Key Concepts

1. The p-value measures the strength of evidence against H_0.
2. We follow a 5-step process to perform a hypothesis test.
3. We compute a **test statistic** (like z) that standardizes the difference between our sample and the null hypothesis.
4. We find the p-value associated with that test statistic.
5. The final decision is a simple comparison: if p-value $\leq \alpha$, reject H_0.
6. Always remember that statistical significance does not automatically imply practical importance.

Exercises

Exercise 1

A machine is supposed to fill bags with 500g of coffee. A quality control inspector takes a sample of 30 bags and finds the average weight is 495g. The population standard deviation is known to be 10g.

- Set up the null and alternative hypotheses to test if the machine is under-filling the bags.
- Calculate the z-statistic.
- If the p-value for this z-statistic is 0.003, what would you conclude at an $\alpha = 0.01$ level?

Exercises

Exercise 1

A machine is supposed to fill bags with 500g of coffee. A quality control inspector takes a sample of 30 bags and finds the average weight is 495g. The population standard deviation is known to be 10g.

- Set up the null and alternative hypotheses to test if the machine is under-filling the bags.
- Calculate the z-statistic.
- If the p-value for this z-statistic is 0.003, what would you conclude at an $\alpha = 0.01$ level?

Exercise 2

Explain why a very large p-value (e.g., 0.95) provides strong support for *not rejecting* the null hypothesis.